深度学习综述

刘静芳

（北京化工大学信息科学技术学院，北京：100029）

摘 要

作为一个十余年来快速发展的崭新领域,深度学习受到了越来越多研究者的关注,它在特征提取和建模上都有着相较于浅层模型显然的优势.深度学习善于从原始输入数据中挖掘越来越抽象的特征表示，而这些表示具有良好的泛化能力.它克服了过去人工智能中被认为难以解决的一些问题。且随着训练数据集数量的显著增长以及芯片处理能力的剧增,它在目标检测和计算机视觉、自然语言处理、语音识别和语义分析等领域成效卓然,因此也促进了人工智能的发展.深度学习是包含多级非线性变换的层级机器学习方法。首先论述了深度学习的基础知识，分析了算法的优越性，并介绍了主流学习算法及应用现状。最后总结了当前存在的问题及发展方向。

摘要：深度学习，CNN, RNN

一、引言

1. 深度学习的定义与背景

机器学习是通过计算模型和算法从数据中学习规律的一门学问，在各种需要从复杂数据中挖掘规律的领域中有很多应用，已成为当今广义的人工智能领域最核心的技术之一。深度学习是机器学习领域一个新的研究方向，近年来在语音识别、计算机视觉等多类应用中取得突破性的进展。深度学习的概念最早由多伦多大学的 G. E. Hinton 等[1]于 2006 年提出，指基于样本数据通过一定的训练方法得到包含多个层级的深度网络结构的机器学习过程。其动机在于建立模型模拟人类大脑的神经连接结构，在处理图像、声音和文本这些信号时，通过多个变换阶段分层对数据特征进行描述，进而给出数据的解释。

人工神经元网络（Artificial Neural Network,ANN）[2]是对生物神经网络的一种模拟和近似，是由大量神经元通过相互连接而构成的自适应非线性动态网络系统。从单层感知器模型的提出到，反向传播 (back-propagation，BP) 算法被应用于训练神经网络，解决了多层感知器无法训练的问题，从而使神经网络具有了非线性表示能力，以 BP 算法训练的多层感知器 (multi-layer perceptron，MLP)[3]成为最成功的神经网络模型。

但神经网络方法也存在很多问题。首先，多层感知器虽然具有极强的非线性表示能力，但也因此导致参数解空间中存在大量的局部极值，使用梯度下降法进行训练很容易产生一个并不好的局部极小值，导致多层感知器在很多问题上推广能力较差。其次，虽然神经网络在理论上可以有很多层，但多层神经网络训练速度很慢，这既是因为当时的硬件条件限制，也是因为多层神经网络存在梯度消散现象，即误差在反向传播过程中会迅速衰减，导致对深层网络权值的修正非常缓慢，因此人们实际上只使用二层或三层的神经网络。对这些问题缺乏如何解决或如何避免的理论指导，实际应用中多靠试算和经验，限制了神经网络的进一步发展，使神经网络研究走向低谷。Hinton 等[4]人基于深信度网(DBN) 提出非监督贪心逐层训练算法，为解决深层结构相关的优化难题带来希望，随后提出多层自

动编码器深层结构。此外 Lecun 等[5]人提出的卷积神经网络(CNN)是第一个真正多层结构学习算法，它利用空间相对关系减少参数数目以提高 BP 训练性能。此外深度学习还出现许多变形结构如去噪自动编码器、DCN 等。

2. 深度学习在 AI 领域的重要性

2012 年的 ImageNet 竞赛中，Krizhevsky 等[6]使用卷积神经网络使准确率提升了 10%，第一次显著地超过了手工设计特征加浅层模型进行学习的模式，在业界掀起了深度学习的热潮。2015 年，Google 旗下 DeepMind 公司研发的 Alpha-Go 使用深度学习方法在围棋比赛中击败了欧洲围棋冠军，使得深度学习影响日益广泛。如今，以 ChatGPT 为代表的生成性人工智能推动着 AI 时代的快速发展。改变了人们的沟通方式：Chat GPT 是一种基于人工智能技术的语言模型，它能够模拟人类的语言和行为，实现人机交互。通过 Chat GPT，人们可以更加便捷地与计算机进行交流，这使得人机交互变得更加自然和流畅。这种交互方式的改变不仅提高了沟通效率，还拓宽了人类的交流渠道。推动了人工智能技术的发展：Chat GPT 的出现是人工智能技术发展的一个里程碑，它不仅展示了人工智能在自然语言处理方面的强大能力，还为后续的人工智能技术发展提供了新的思路和方法。通过 Chat GPT，人们可以更加深入地了解人工智能技术的原理和应用，从而推动人工智能技术的进一步发展。改变了信息获取和传递的方式：Chat GPT 可以自动学习和理解人类的语言，通过自然语言处理技术对信息进行分类、分析和提取，从而为用户提供更加精准和高效的信息服务。这使得信息获取和传递的方式发生了重大变革，人们可以更加便捷地获取自己需要的信息，提高了信息利用的效率和准确性。Chat GPT 的出现对当今社会产生了广泛而深远的影响，推动了人机交互、人工智能技术、信息获取和传递等方面的变革。人们对人工智能的关注达到了前所未有的程度。

二、深度学习基本原理

深度学习最常用于各种监督模式识别问题，比如图像识别、自然语言识别等。在讨论深度学习的典型模型之前，我们先来讨论作为各种深度学习模型和算法共同基础的核心学习算法。一般地，深度神经网络包含输入层、多个隐含层以及输出层，传统多层感知器神经网络训练的反向传播 (BP) 算法仍然是深度神经网络

训练的核心算法，它包括信息的前向传播过程和误差梯度的反向传播过程[8]。

2.1 神经元

神经元是人工神经网络的基本处理单元，神经元的 M-P 模型如图所示：

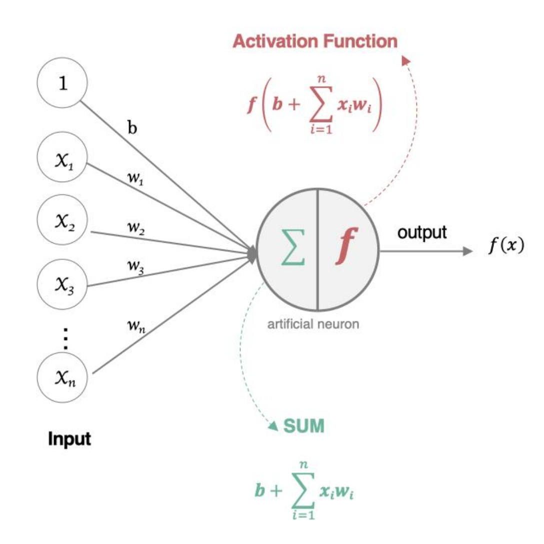


图 1 神经元的 M-P 结构图

图中，$x_i,(i=1,2,...,n)$ 表示神经元的输入，$w_i(i=1,2,...,n)$ 表示输入信号与连接神经元之间的权重值，b 表示偏置，神经元的输出可表示为：

$$y=f(\sum_{i=1}^{n} w_i x_i + b)$$

$f(\cdot)$ 表示激活函数，其可以有多种选择，sigmiod 函数，tanh(x) 函数，径向基函数等。

2. 2 多层感知机

多层感知器的基本结构如图 2 所示，多层感知器有前向传播与反向传播两个过程。

输入层神经元接收输入信号，隐含层和输出层的每一个神经元与之相邻层的所有神经元连接，即全连接，同一层的神经元间不相连。图 2 中，有箭头的线段表示神经元间的连接和信号传输的方向，且每个连接都有一个连接权值.隐含层和输出层中每一个神经元的输入为前一层所有神经元输出值的加权和。

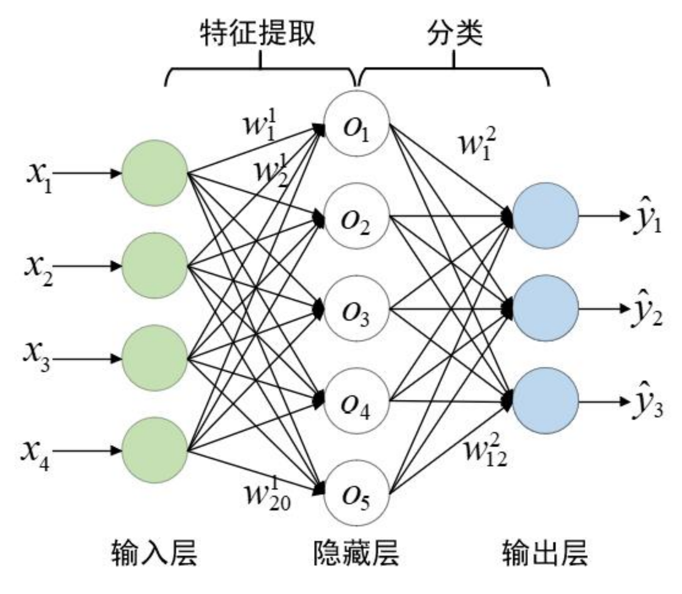


图 2 多层感知器的结构

信号在网络中前向传播的过程,每个节点中都包含 2 步操作,先对上一层节点输出值进行线性组合，再对得到的中间值进行非线性变换后输出。对于 1 个输入样本,经过上述 2 步操作可以得到第 1 层隐含节点的输出值，隐含节点输出值就是特征的某种抽象表示，可以重复这个过程得到更深层次的隐含节点值，越深层次的隐含节点所表示的特征越抽象，对于最后一层隐含节点，可以连接到输出层中进行分类并输出。

当输出结果与真实标签相等时损失为零，二者相差越大损失函数值越大，常见的损失函数有二次损失、对数损失等。在训练样本上的总损失是监督学习中的优化目标，常用梯度下降法优化这个目标，这个过程就是机器的“学习”或用样本对机器的“训练”[9]。

2.3 反向传播算法

要对神经网络各层的参数进行训练,需要计算损失对网络中间各层参数的梯度，BP 算法就是把损失从输出层逐层往前传递，这个过程叫做误差的反向传播。算法的核心是用链式求导法从输出层逐层向前计算损失函数对隐含节点输出值的梯度和对连接权重的梯度。将连接权重向负梯度方向适度调整得到新一轮的参数。用大量样本如此循环训练多次，直到损失函数不再下降或达到设定的迭代次数,就完成了神经网络的训练过程。

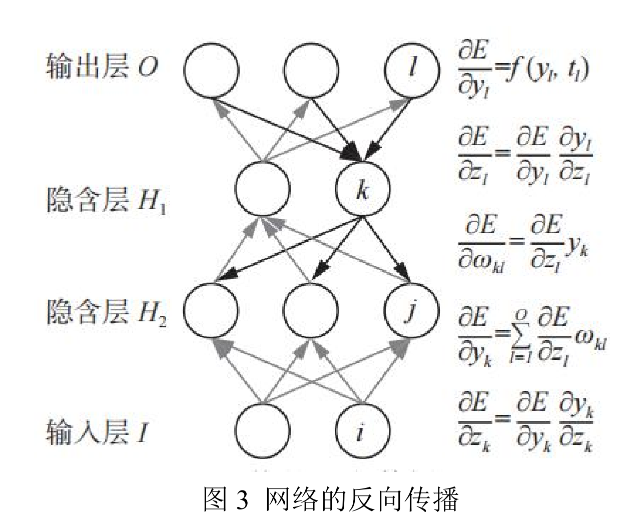


图 3 网络的反向传播

2.4 激活函数与非线性变换

激活函数（Activation Function）[10]是一种添加到人工神经网络中的函数，旨在帮助网络学习数据中的复杂模式。在神经元中，输入的 input 经过一系列加权求和后作用于另一个函数，这个函数就是这里的激活函数。激活函数可以分为线性激活函数（线性方程控制输入到输出的映射，如 f(x)=x 等）以及非线性激活函数（非线性方程控制输入到输出的映射，比如 Sigmoid、Tanh、ReLU、LReLU[11]等）。一般来说，在神经元中，激活函数是很重要的一部分，为了增强网络的表示能力和学习能力，神经网络的激活函数都是非线性的，通常具有以下几点性质：连续并可导（允许少数点上不可导），可导的激活函数可以直接利用数值优化的方法来学习网络参数；激活函数及其导数要尽可能简单一些，太复杂不利于提高网络计算率；激活函数的导函数值域要在一个合适的区间内，不能太大也不能太小，否则会影响训练的效率和稳定性。常用的激活函数有

2.4.1 Sigmoid 函数

Sigmoid[12]函数也叫 Logistic 函数，用于隐层神经元输出，取值范围为(0,1)，它可以将一个实数映射到(0,1)的区间，可以用来做二分类。在特征相差比较复杂或是相差不是特别大时效果比较好。Sigmoid 是一个十分常见的激活函数，函数的表达式如下：

$$f(x)=\frac{1}{1+e^{-x}}$$

图像如下：

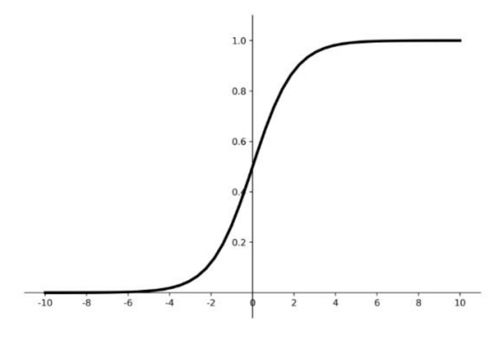


图 4　Sigmoid 函数图像

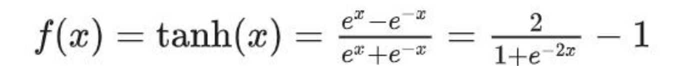
Sigmoid 函数的输出范围是 0 到 1。由于输出值限定在 0 到 1，因此它对每个神经元的输出进行了归一化；适用于将预测概率作为输出的模型。由于概率的取值范围是 0 到 1，因此 Sigmoid 函数非常合适；梯度平滑，避免跳跃的输出值；函数是可微的。这意味着可以找到任意两个点的 Sigmoid 曲线的斜率；明确的预测，即非常接近 1 或 0。

Sigmoid 激活函数存在的不足：存在梯度消失现象。Sigmoid 函数趋近 0 和 1 的时候变化率会变得平坦，也就是说，Sigmoid 的梯度趋近于 0。神经网络使用 Sigmoid 激活函数进行反向传播时，输出接近 0 或 1 的神经元其梯度趋近于 0。这些神经元叫作饱和神经元。因此，这些神经元的权重不会更新。此外，与此类神经元相连的神经元的权重也更新得很慢。该问题叫作梯度消失。因此，如果一个大型神经网络包含 Sigmoid 神经元，而其中很多个都处于饱和状态，那么该网络无法执行反向传播。不以零为中心，Sigmoid 输出不以零为中心的,，输出恒大于 0，非零中心化的输出会使得其后一层的神经元的输入发生偏置偏移（Bias Shift），并进一步使得梯度下降的收敛速度变慢。计算成本高昂，exp()函数与其他非线性激活函数相比，计算成本高昂，计算机运行起来速度较慢。

2.4.2 Tanh 双曲正切函数

Tanh 激活函数又叫作双曲正切激活函数（hyperbolic tangent activation function）。与 Sigmoid 函数类似，Tanh 函数也使用真值，但 Tanh 函数将其压缩至-1 到 1 的区间内。与 Sigmoid 不同，Tanh 函数的输出以零为中心，因为区间在-1 到 1 之间。

函数表达式为：



$$f(x)=\tanh(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}=\frac{2}{1+e^{-2x}}-1$$

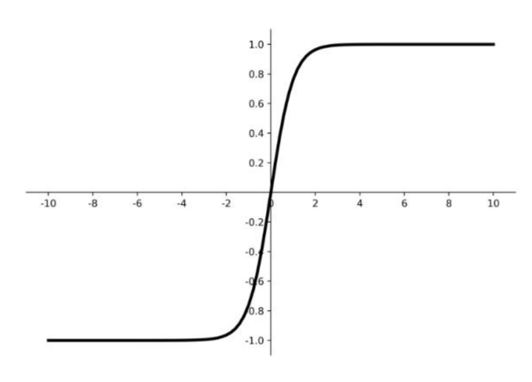


图 5 tanh 函数图像

3.4.3 ReLU 激活函数

ReLU 函数又称为修正线性单元（Rectified Linear Unit），是一种分段线性函数，其弥补了 sigmoid 函数以及 tanh 函数的梯度消失问题，在目前的深度神经网络中被广泛使用。ReLU 函数本质上是一个斜坡（ramp）函数，公式及函数图像如下：

$$f(x)=\begin{cases}x & ,x>=0\\ 0 & ,x<0\end{cases}$$

ReLU 激活函数的提出 就是为了解决梯度消失问题，LSTMs 也可用于解决梯度消失问题(但仅限于 RNN 模型)。ReLU 的梯度只可以取两个值：0 或 1，当输入小于 0 时，梯度为 0；当输入大于 0 时，梯度为 1。好处就是：ReLU 的梯度的连乘不会收敛到 0 ，连乘的结果也只可以取两个值：0 或 1 ，如果值为 1 ，梯度保持值不变进行前向传播；如果值为 0 ,梯度从该位置停止前向传播。

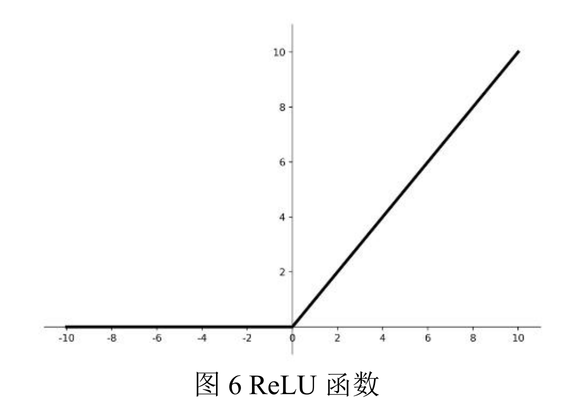


图 6 ReLU 函数

三、深度学习的主要技术

3.1 卷积神经网络（CNN）

CNN [13]的基本结构由输入层、卷积层(convolutional layer)、池化层(pooling layer，也称为取样层)、全连接层及输出层构成.卷积层和池化层一般会取若干个，采用卷积层和池化层交替设置，即一个卷积层连接一个池化层，池化层后再连接一个卷积层依此类推.由于卷积层中输出特征面的每个神经元与其输入进行局部连接，并通过对应的连接权值与局部输入进行加权求和再加上偏置值，得到该神经元输入值，该过程等同于卷积过程，CNN 也由此而得名。

3.1.1 卷积层的工作原理

卷积层由多个特征面(Feature Map)组成，每个特征面由多个神经元组成，它的每一个神经元通过卷积核与上一层特征面的局部区域相连。卷积核是个权值矩阵(如对于二维图像而言可为 3 * 3 或 5*5 矩阵)。CNN 的卷积层通过卷积操作提取输入的不同特征，第 1 层卷积层提取低级特征如边缘、线条、角落，更高层的卷积层提取更高级的特征。

3.1.2 池化层的作用

池化层紧跟在卷积层之后，同样由多个特征面组成，它的每一个特征面唯一对应于其上一层的一个特征面，不会改变特征面的个数.卷积层是池化层的输入层，卷积层的一个特征面与池化层中的一个特征面唯一对应，且池化层的神经元也与其输入层的局部接受域相连，不同神经元局部接受域不重叠.池化层旨在通过降低特征面的分辨率来获得具有空间不变性的特征.池化层起到二次提取特征的作用，它的每个神经元对局部接受域进行池化操作.常用的池化方法有最大池化即取局部接受域中值最大的点、均值池化即对局部接受域中的所有值求均值、随机池化。Boureau 等[14]人给出了关于最大池化和均值池化详细的理论分析，通过分析得出以下一些预测：(1)最大池化特别 适用于分离非常稀疏的特征；(2) 使用局部区域内所有的采样点去执行池化操作也许不是最优的，例如均值池化就利用了局部接受域内的所有采样点。Boureau 等[15]人比较了最大池化和均值池化两

种方法，通过实验发现：当分类层采用线性分类器如线性 SVM 时，最大池化方法比均值池化能够获得一个更好的分类性能。

3.1.3 全连接层

在 CNN 结构中，经多个卷积层和池化层后，连接着 1 个或 1 个以上的全连接层与 MLP 类似，全连接层中的每个神经元与其前一层的所有神经元进行全连接。全连接层可以整合卷积层或者池化层中具有类别区分性的局部信息。为了提升 CNN 网终性能，全连接层每个神经元的激励函数一般采用 ReLU 函数。最后一层全连接层的输出值被传递给一个输出层，可以采用 softmax 逻辑回归(softmax-regression)进行分类，该层也可称为 softmax 层(softmax layer)。

3.2 循环神经网络（RNN）与长短期记忆网络（LSTM）

循环神经网络(RNN) [16]是一类非常强大的用于处理和预测序列数据的神经网络模型。循环结构的神经网络克服了传统机器学习方法对输入和输出数据的许多限制，使其成为深度学习领域中一类非常重要的模型。RNN 及其变体网络已经被成功应用于多种任务，尤其是当数据中存在一定时间依赖性的时候。语音识别、机器翻译、语言模型、文本分类、词向量生成、信息检索等，都需要一个模型能够将具有序列性质的数据作为输入进行学习。RNN 通常难以训练，循环多次之后，大多数情况下梯度往往倾向于消失，也有较少情况会发生梯度爆炸的问题。针对 RNN 在实际应用中存在的问题，长短期记忆(LSTM) 网络被提出，它能够保持信息的长期存储而备受关注。

3.2.1RNN 的工作机制

RNN 是深度学习领域中一类特殊的内部存在自连接的神经网络,可以学习复杂的矢量到矢量的映射。RNN 的网络结构如图：

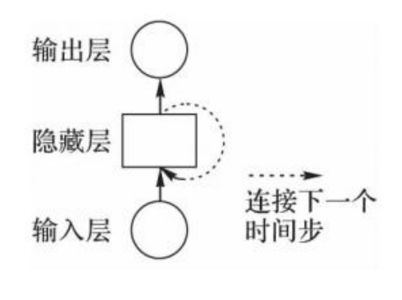


图 7 RNN 网络结构

通过隐藏层上的回路接，使得前一时刻的网络状态能够传递给当前时刻，当前时刻的状态也可以传递给下个时刻。可以将 RNN 看作所有层共享权值的深度 FNN,通过连接两个时间步来扩展。参数共享的概念早在隐马尔可夫模型Hidden Markov Model，HMM) 中就已经出现，HMM 也常用于京列数据建模并且在语音识别领域一度取得很好的效果。HMM 和 RNN 均使用内部状态来表示序列中的依赖关系。当时间序列数据存在长距离的依赖，并且该依赖的范围随时间变化或者未知，那么 RNN 可能是相对较好的解决方案。展开后的 RNN 结构如下：

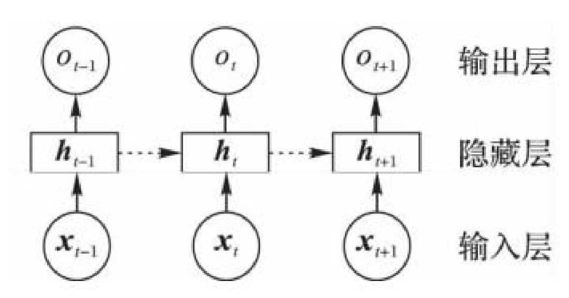


图 8 RNN 展开后的结构

对于 RNN 的输入和输出，下图中:(a)表示传统的、固定尺度的输入到固定尺度的输出；(b)序列输入，可用于表示例如情感分析等任务，给定句子然后将其与一个情感表示向量关联；(c)序列输出，可以用于表示图片标注等任务，输入固定大小的向量表示的图片输出图片描述；(d)和(e)中的输入和输出均为序列数据，目输入和输出分别为非同步和同步，(d)可以用于机器翻译等任务，(e)常用于语音识别中。

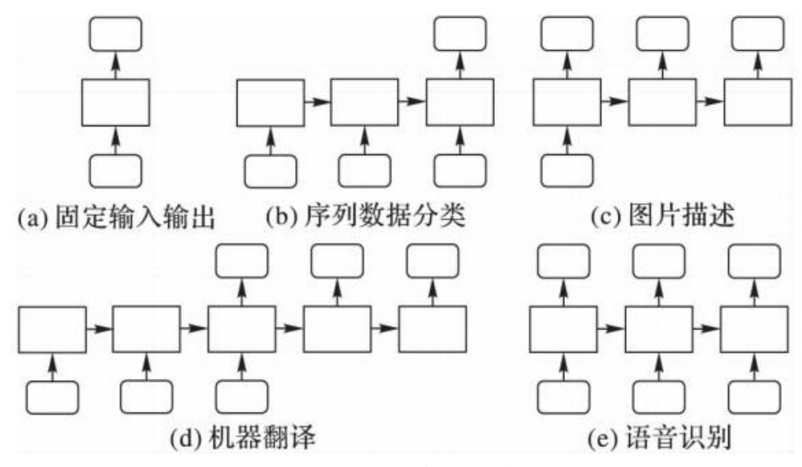


图 9 RNN 的输入和输出

实际应用中，RNN 常常面临训练方面的难题;尤其随着模型深度不断增加，使得 RNN 并不能很好地处理长距离的依赖。Jacobian 矩阵的乘积往往会以指数级增大或者减小,其结果是使得长期依赖特别困难。RNN 在反向传播时，由于参

数共享和多次连乘的特性，容易出现梯度消失或梯度爆炸的问题，导致模型难以训练或无法收敛。

3.2.2 LSTM 如何解决 RNN 的梯度消失问题

目前，在实际应用中使用最广泛的循环结构网络架构是 LSTM(Long Short-Term Memory)[17]，它能够有效克服 RNN 中存在的梯度消失问题，尤其在长距离依赖的任务中的表现远优于 RNNG,梯度反向传播过程中不会再受到梯度消失问题的困扰，可以对存在短期或者长期依赖的数据进行精确的建模。LSTM 的工作方式与 RNN 基本相同区别在于 LSTM 实现了一个更加细化的内部处理单元，来实现上下文信息的有效存储和更新。

LSTM 单元结构如下：

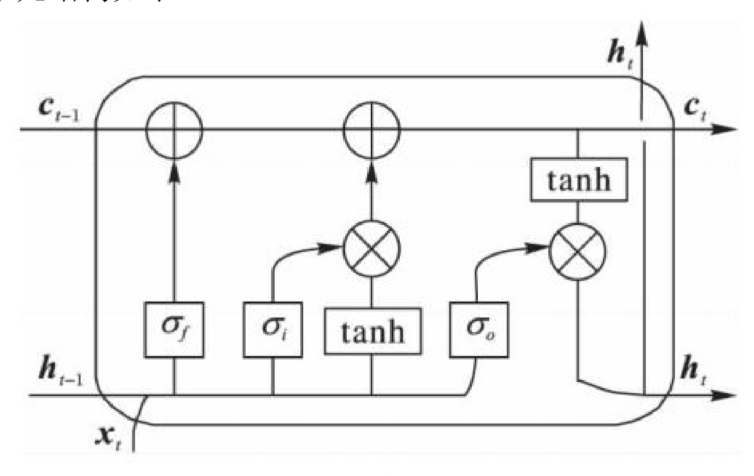


图 10 LSTM 单元结构如下

LSTM 单元中有三种类型的门控，分别为:输入门、遗忘门和输出门。门控可以看作一层全连接层,ISTM 对信息的存储和更新正是由这些门控来实现。更具体地说，门控是由 sigmoid 函数和点乘运算实现，门控并不会提供额外的信息。门控的一般形式可以表示为:

$$g(x) = \sigma(Wx + b)$$

LSTM 的计算过程如下：

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$

$$h_t = o_t \circ \tanh(c_t)$$

遗忘门是 LSTM 单元的关键组成部分，可以控制哪些信息要保留哪些要遗忘，并且以某种方式避免当梯度随时间反向传播时引发的梯度消失和爆炸问题。遗忘门控制自连接单元，可以决定历史信息中的哪些部分会被丢弃。即上一时刻记忆单元c_{t-1}中的信息对当前记忆单元c_t的影响。

3.3 生成对抗网络（GAN）

生成式对抗网络 GAN (Generative adversarial networks)[18]目前已经成为人工智能学界一个热门的研究方向。GAN 的基本思想源自博弈论的二人零和博弈，即二人的利益之和为零, 一方的所得正是另一方的所失。由一个生成器和一个判别器构成，生成器捕捉真实数据样本的潜在分布, 并生成新的数据样本; 判别器是一个二分类器, 判别输入是真实数据还是生成的样本. 生成器和判别器均可以采用目前研究火热的深度神经网络，通过对抗学习的方式来训练。目的是估测数据样本的潜在分布并生成新的数据样本。优化过程是一个极小极大博弈(Minimax game) 问题, 优化目标是达到纳什均衡, 使生成器估测到数据样本的分布。

3.3.1 GAN 的基本结构和工作原理

GAN 的核心思想来源于博弈论的纳什均衡。它设定参与游戏双方分别为一个生成器 (Generator 和一个判别器(Discriminator)，生成器的目的是尽量去学习真实的数据分布，而判别器的目的是尽量正确判别输入数据是来自真实数据还是来自生成器为了取得游戏胜利，这两个游戏参与者需要不断优化，各自提高自己的生成能力和判别能力，这个学习优化过程就是寻找二者之间的一个纳什均衡.GAN 的计算流程与结构如图所示. 任意可微分的函数都可以用来表示 GAN 的生成器和判别器，由此我们用可微分函数 D 和 G 来分别表示判别器和生成器，它们的输入分别为真实数据 x 和随机变量 z。

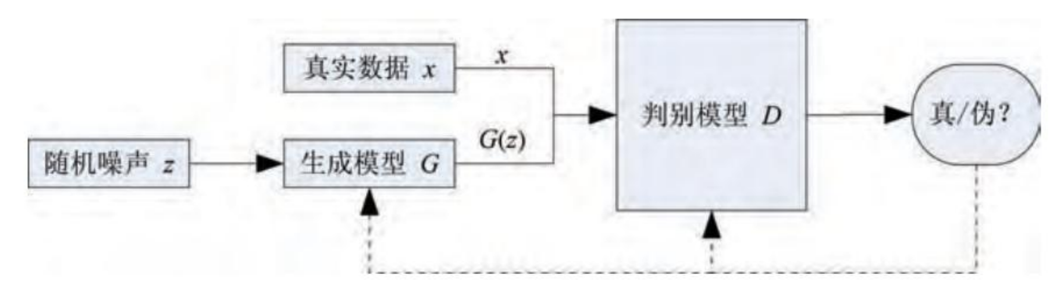


图 11 GNN 原理图

G(z)则为由 G 生成的尽量服从真实数据分布 pdata 的样本，如果判别器的输入来自真实数据，标注为 1。如果输入样本为 G(z)，标注为 0。这里 D 的目标是实现对数据来源的二分类判别：真(来源于真实数据 x 的分布) 或者伪(来源于生成器的伪数据 G(z)。而 G 的目标是使自己生成的伪数据 G(z)在 D 上的表现 D(G(z))和真实数据 a 在 D 上的表现 D(a）一致，这两个相互对抗并迭代优化的过程使得 D 和 G 的性能不断提升，当最终 D 的判别能力提升到定程度，并且无法正确判别数据来源时，可以认为这个生成器 G 已经学到了真实数据的分布。

对于 GAN 的学习过程，我们需要训练模型 D 来最大化判别数据来源于真实数据或者伪数据分布 G(z)的准确率，同时，我们需要训练模型 G 来最小化 log(1 - D(G(z))).这里可以采用交优化的方法： 先固定生成器 G，优化判别器 D，使得 D 的判别准确率最大化；然后固定判别器 D，优化生成器 G，使得 D 的判别准确率最小化。当且仅当 Pdata = p 时，达到全局最优解。训练 GAN 时，同一轮参数更新中，一般对 D 的参数更新 k 次再对 G 的参数更新 1 次。

GAN 对于生成式模型的发展具有重要的意义，GAN 作为一种生成式方法，有效解决了可建立自然性解释的数据的生成难题。尤其对于生成高维数据所采用的神经网络结构不限制生成维度，大大拓宽了生成数据样本的范围.所采用的神经网络结构能够整合各类损失函数，增加了设计的自由度。

四、未来展望与研究方向

深度学习已成功应用于多种模式分类问题。这一领域虽处于发展初期,但它的发展无疑会对机器学习和人工智能系统立生影响。同时它仍存在某些不适合处理的特定任务，譬如语言辨识,生成性预训练提取的特征仅能描述潜在的语音变化不会包含足够的不同语言间的区分性信息;虹膜识别等每类样本仅含单个样本的模式分类问题也是不能很好完成的任务。

深度学习目前仍有大量工作需要研究。模型方面是否有其他更为有效且有理论依据的深度模型学习算法,探索新的特征提取模型是值得深入研究的内容。此外有效的可并行训练算法也是值得研究的一个方向。当前基于最小批处理的随机梯度优化算法很难在多计算机中进行并行训练。通常办法是利用图形处理单元加速学习过程,然而单个机器 GPU 对大规模数据识别或相似任务数据集并不适用。在深度学习应用拓展方面,如何充分合理地利用深度学习在增强传统学习算法的性能仍是目前各领域的研究重点。

参考文献

[1] 高隽.人工神经网络原理及仿真实例[M].北京：机械工业出版社,2003.7.

[2] McCulloch W S，Pitts W.　A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics，1943，5(4)：115-133.

[3] LeCun Y，Bottou L，Bengio Y，et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE，1998，86(11)：2278-2324.

[4] 高君宇，杨小汕，张天柱等. 基于深度学习的鲁棒性视觉跟踪方法. 计算机学报，2016，39（7）：1419-1432.

[5] 周飞燕，金林鹏，董军. 基于集成学习的室性早博识别方法. 电子学报，2017，45（2）：501-507.

[6] MCMAHAN H B, HOLT G, SCULLEY D, et al. Ad click prediction: a view from the trenches[C]//Proceedings of the19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Chicago, USA, 2013：1222-1230.

[7] 周永进.BP 网络的改进及其应用[D].南京：南京信息工程大学，2007.

[8] 张立明. 人工神经网络的模型及其应用. 上海：复旦大学出版社，1993.

[9] 阎平凡，张长水. 人 工 神 经 网 络 与 模 拟 进 化 计 算. 北 京：清华大学出版社，2005.

[10] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1)：32-42.

[11] Van HAMME D, VEELAERT P, PHILIPS W. Robust visual odometry using uncertainty models [C] // Proc of the 13th International Conference on Advanced Concepts for Intelligent Vision Systems. Berlin: Springer-Verlag ,2011：1-12.

[12] 焦李成，周伟达，张莉等. 智能目标识别与分类. 北京：科学出版社，2010.

[13] 虞和济. 基于神经网络的智能诊断. 北 京：冶 金 工 业 出 版社，2002.

[14] GUIZILINI V，RAMOS F. Visual odometry learning for unmanned aerial vehicles [C] //Proc of IEEE International Conference on Robotics and Automation. 2011：6213-6220.

[15] 杨行峻. 人工神经网络与盲信号处理. 北 京：清 华 大 学 出版社，2003.

[16] GREFF K，SRIVASTAVA R K，KOUTNIK J，et al. LSTM：a search space Odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems，2017，28

(10)：2222 - 2232.

[17] 王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. 自动化学报, 2016, 42(10): 1490−1500.

[18] Yu L T, Zhang W N, Wang J, Yu Y. SeqGAN: sequence generative adversarial nets with policy gradient. arXiv preprint arXiv: 1609. 05473，2016.